



Classical ensembles of single-qubit quantum variational circuits for classification

Shane McFarthing^{1,2} · Anban Pillay^{1,3} · Ilya Sinayskiy^{2,4} · Francesco Petruccione^{2,5}

Received: 22 February 2023 / Accepted: 25 October 2024
© The Author(s) 2024

Abstract

The quantum asymptotically universal multi-feature (QAUM) encoding architecture was recently introduced and showed improved expressivity and performance in classifying pulsar stars. The circuit uses generalized trainable layers of parameterized single-qubit rotation gates and single-qubit feature encoding gates. Although the improvement in classification accuracy is promising, the single-qubit nature of this architecture, combined with the circuit depth required for accuracy, limits its applications on NISQ devices due to their low coherence times. This work reports on the design, implementation, and evaluation of ensembles of single-qubit QAUM classifiers using classical bagging and boosting techniques. We demonstrate an improvement in validation accuracy for pulsar star classification. We find that this improvement is not dataset specific as we observe consistent improvements for the MNIST Digits and Wisconsin Cancer datasets. We also observe that the boosting ensemble achieves an acceptable level of accuracy with only a small amount of training, while the bagging ensemble achieves higher overall accuracy with ample training time. This shows that classical ensembles of single-qubit circuits present a new approach for certain classification problems.

Keywords Quantum computing · Machine learning · Quantum classifiers · Classical ensembles

1 Introduction

The potential to efficiently run classically intractable algorithms on quantum devices has driven the development of both quantum algorithms and quantum computers them-

selves, with the goal of obtaining a fully fledged quantum device with error-mitigation (Nielsen and Chuang 2002). Quantum computers utilize the principles of quantum mechanics to operate on data and are built on the foundational concept of the qubit, a quantum bit which, unlike the classical bit, can also exist in a superposition of the states 0 and 1 (Nielsen and Chuang 2002). Additionally, quantum entanglement allows the qubits in a quantum computer to be correlated in a manner not possible in classical computing, resulting in a significant increase in computational power due to the possibility of parallel operations on multiple qubits (Nielsen and Chuang 2002).

The realization of full-scale quantum computers on which it is possible to perform error-correction to deal with noise is a significant engineering challenge (Schuld and Petruccione 2018b). Implementing robust error-correction schemes requires quantum devices possessing many orders of magnitude more qubits than the quantum computers available today (Schuld and Petruccione 2018b). As a result, noisy intermediate-scale quantum (NISQ) devices present a viable platform for the development of new quantum algorithms and techniques for mitigating errors due to noise (Nielsen and Chuang 2002). These NISQ devices have much lower

✉ Shane McFarthing
shane.mcfarthing@gmail.com

Ilya Sinayskiy
sinayskiy@ukzn.ac.za

¹ School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal (Westville Campus), Durban 4001, South Africa

² National Institute for Theoretical and Computational Sciences (NITheCS), Stellenbosch 7600, South Africa

³ Centre for Artificial Intelligence Research (CAIR), Cape Town, South Africa

⁴ School of Chemistry and Physics, University of KwaZulu-Natal (Westville Campus), Durban 4001, South Africa

⁵ School of Data Science and Computational Thinking and Department of Physics, Stellenbosch University, Cape Town 7600, South Africa

numbers of qubits and cannot run error-correction routines to combat noise; however, they are still capable of performing tasks that are difficult on classical computers, such as the simulation of quantum systems (Schuld and Petruccione 2018b).

Many near-term algorithms have been developed to run on these NISQ devices as a way to leverage the hardware available today. One particular class of algorithms that can efficiently utilize NISQ devices are hybrid quantum-classical algorithms (Schuld and Petruccione 2018b). These algorithms prepare and run parameterized quantum circuits (PQCs) on quantum computers and use the output to optimize the parameters of the PQC on classical computers (Nielsen and Chuang 2002). Some examples are the quantum approximate optimization algorithm (QAOA), which can be applied to combinatorial optimization problems, and the variational quantum eigensolver (VQE) algorithm for computing the ground state energy of a given Hamiltonian, with applications in nuclear physics and nuclear structure problems among others (Nielsen and Chuang 2002).

The field of quantum machine learning (QML) resulted from the application of quantum computers to solve machine learning problems. QML involves the construction and training of mathematical models capable of learning patterns and relationships in data and then applying these to new data (Schuld and Petruccione 2018b). Many machine learning algorithms obtain the best solution to a problem using optimization, and quantum machine learning makes use of the unique properties of quantum systems to speed-up this optimization procedure (Schuld and Petruccione 2018b). Despite the relatively recent emergence of this field, numerous quantum machine learning algorithms have already been devised, such as quantum neural networks (QNN) and quantum support vector machines (QSVM), and have shown promise in various classical machine learning problems (Schuld and Petruccione 2018b).

These quantum machine learning algorithms are mostly of a hybrid quantum-classical nature and make extensive use of PQCs (Schuld and Petruccione 2018b). The choice of circuit structure used in these algorithms is very important as it determines both the expressivity of the model (Schuld et al. 2021) and its trainability (McClean et al. 2018). As a result, the creation of PQCs with novel architectures capable of achieving high levels of generalization is an ongoing area of development (Schuld et al. 2015).

One such novel architecture is the quantum asymptotically universal multi-feature (QAUM) encoding single-qubit circuit explored in Kordzanganeh et al. (2021). This circuit design was inspired by Schuld et al. (2021), which showed that a quantum model can be expressed as a partial Fourier

series in the data, and that as more Fourier coefficients are made accessible through repetition of data encoding gates, the expressivity of the quantum model increases. The QAUM circuit makes use of generalized trainable layers of parameterized single-qubit rotation gates and single-qubit feature encoding gates and was used for the classification of pulsar stars (Kordzanganeh et al. 2021). When compared with the QAOA, the model built with the QAUM single-qubit circuit demonstrated the ability to obtain higher accuracy in its classification of the HTRU 2 Pulsar dataset (Kordzanganeh et al. 2021; Lyon et al. 2016). Although this improvement in classification accuracy is promising, the single-qubit nature of this architecture and the circuit depth required for accuracy limit its applications on NISQ devices due to their low coherence times (Kordzanganeh et al. 2021).

In this work, we explore the application of classical ensemble techniques with the QAUM classifier as means to scale up the QAUM architecture. We do this by utilizing multiple single-qubit QAUM circuits combined in an ensemble, rather than expanding the design of the QAUM circuit to multiple qubits. Though the concept of using quantum ensembles is not new, with Schuld and Petruccione (2018a); Macaluso et al. (2020) detailing techniques for creating ensembles of quantum states, the creation of classical ensembles of variational quantum classifiers is relatively unexplored. Ensembles of variational quantum classifiers that use classical voting strategies are explored in Qin et al. (2022) and were shown to produce valuable improvements in classification accuracy over comparable quantum classifiers on IBM's NISQ devices. Classical bagging and boosting ensembles of quantum variational classifiers have also been explored in Li et al. (2023), which showed that higher accuracies and increased robustness are attainable using these ensemble techniques. Compared to wholly classical ensembles, quantum ensembles of quantum classifiers have been shown to have favorable scaling with regards to both ensemble size and training time. However, the number of qubits required to implement these quantum ensembles is currently out of reach of current-era quantum devices (Schuld and Petruccione 2018a; Macaluso et al. 2020).

The ensembles explored in this work utilize single-qubit circuits which can be executed on NISQ devices. Unlike Qin et al. (2022), which uses plurality voting that is susceptible to classification errors on outlying datapoints, and Li et al. (2023), which uses majority voting, we utilize a metalearner to obtain the optimal weighted voting strategy. Although both works (Qin et al. 2022; Li et al. 2023) also investigate the use of classical ensemble techniques, they utilize generalized multi-qubit circuits, whereas we have specifically chosen the QAUM single-qubit circuit from Kordzanganeh et al. (2021).

This was done as our goal is to investigate the scaling of single-qubit architectures using ensemble techniques, rather than the more general use of ensemble techniques in Qin et al. (2022); Li et al. (2023). To the best of our knowledge, our work is the first to explore this particular relationship.

The single-qubit nature of the classifiers in our work makes the use of distributed quantum computing resources a possibility, as well as lower the requirements needed to run the ensembles on NISQ hardware. The ensembles use the QAUM classifier as their base learners, and we test their classification performance on the HTRU 2 Pulsar dataset (Lyon et al. 2016). We find that the boosting and bagging ensembles outperform the single QAUM classifier from Kordzanganeh et al. (2021) when the amount of training permitted is both severely restricted and virtually unrestricted, respectively. We performed further testing on the MNIST Digits (LeCun and Cortes 2010) and Wisconsin Breast Cancer (Dua and Graff 2017) datasets and verified that these improvements in performance were not limited to the pulsar classification problem.

Additionally, although many algorithms involving the variational training of parameterized quantum circuits suffer from exponentially vanishing gradients, known as the barren plateau problem (McClean et al. 2018), this is very unlikely for our methods. The high-dimensional parameter space of the quantum circuits and high levels of entanglement are some of the main causes of the barren plateau phenomenon (Ortiz Marrero et al. 2021), and as we use only single-qubit parameterized circuits, both of these qualities are minimized. This can form a trade-off as single-qubit circuits are not as powerful as higher-dimensional circuits with large levels of entanglement, but avoid barren plateaus and can be used to create ensembles that are more powerful than the individual single-qubit classifiers.

The rest of this paper is organized as follows: in Section 2, we briefly introduce the formalism of classification problems as well as the foundational work we build upon; in Section 3, we detail the methodology used for the construction and testing of the ensembles; in Section 4, we present the results obtained from the testing of the ensembles; and in Section 5, we give our conclusions and notes for areas of future work. Additionally, the graphs containing the results of each run of the cross-validation may be found in Appendices A and B.

2 Background information

In machine learning, an n -dimensional binary classification problem can be described as a search for a model f with i parameters which maps n -dimensional data to one of two

classes, as shown below:

$$f_{\theta}(x) = y, \quad (1)$$

where $\theta = [\theta_1, \theta_2, \dots, \theta_i]$ are the parameters of the model, $x \in \mathbb{R}^n$ is the input data, and $y \in \{0, 1\}$ is the class label assigned by the model, with 0 and 1 being the binary class options.

By encoding the data and parameters of a classification problem into quantum states and using quantum gates to operate on these states, we are able to create quantum classification models (Schuld and Petruccione 2018b). The classification of the input can be obtained by measuring the output of these PQCs, often referred to as quantum neural networks.

2.1 Quantum asymptotically universal multi-feature (QAUM) architecture

The QAUM single-qubit model comprises two circuit blocks: trainable layers made of parameterized single-qubit gates and the multi-feature encoding gates (Kordzanganeh et al. 2021). For the encoding of the features of the dataset, each feature is scaled to lie in the range $(0, \pi)$ and used as a parameter for a Z -rotation. Given an n -dimensional feature vector, a single layer of the QAUM circuit contains n encoding gates and $n+1$ trainable layers which are situated in an alternating fashion.

The expressivity of this architecture, by which we mean the classes of functions that it can learn, is a result of its capability to express the first degree multi-dimensional Fourier series of the dataset used (Kordzanganeh et al. 2021). All of the equations and derivations below have been taken from Kordzanganeh et al. (2021) and show that a truncated Fourier series of higher frequency can be obtained by increasing the number of repetitions in the circuit:

A circuit with a single encoding can be expressed in tensor notation as

$$|\psi\rangle = W^{(2)} S(x) W^{(1)} |0\rangle \rightarrow W_{ki}^{(2)} e^{i\mathcal{G}_{ij}x} W_{j1}^{(1)}, \quad (2)$$

where $S(x)$ is the encoding used and W are the weights of the trainable layers. In this case, we specify $S(x) = e^{i\mathcal{G}_{ij}x}$, with \mathcal{G} being the Pauli-Z matrix. Additionally, $|0\rangle \rightarrow [1, 0]^T$ is absorbed into the second index of W^1 , and the Einstein summation convention is used.

Generalizing to the case with N features and L encoding repetitions, and using the eigenvalues of \mathcal{G} , $\lambda \in \{-1, 1\}$, we

can rewrite (2) as

$$|\psi\rangle_k \rightarrow W_{ki_{L1}}^{(L^1)} e^{i\lambda_{i_{L1}x_1}} W_{i_{L1}i_{L2}}^{(L^2)} \dots W_{i_{LN-1}i_{LN}}^{(L^N)} e^{i\lambda_{i_{LN}x_N}} W_{i_{LN}i_{(L-1)1}}^{((L-1)^N)} \dots W_{i_{1N}1}^{(0)}. \quad (3)$$

The expectation value of a measurement operator, \hat{M} , in the $|\psi\rangle$ basis is obtained through a measurement of this qubit in the basis defined by \hat{M} and is

$$\langle\psi|\hat{M}|\psi\rangle = W_{1j_{1N}}^{\dagger(0)} \dots W_{j_{L1}k'}^{\dagger(L^1)} M_{k'k} W_{ki_{L1}}^{(L^1)} \dots W_{i_{1N}1}^{(0)} e^{i \sum_{l=1}^N x_l \left(\sum_{m=1}^L \lambda_{i_{ml}} - \sum_{p=1}^L \lambda_{j_{pl}} \right)}. \quad (4)$$

We can rewrite the exponent in Eq. 4 as

$$\begin{aligned} \beta &= \sum_{l=1}^N x_l \left(\sum_{m=1}^L \lambda_{i_{ml}} - \sum_{p=1}^L \lambda_{j_{pl}} \right) \\ &= [1, 1, \dots, 1] \left([\lambda_{i_{ml}}] - [\lambda_{j_{pl}}] \right) [x_1, x_2, \dots, x_N]^T. \end{aligned} \quad (5)$$

Through further simplification, we obtain

$$\beta = 2\gamma [x_1, x_2, \dots, x_N]^T, \quad (6)$$

where $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_N]$, with $\gamma_q \in \{-L, -(L-1), \dots, -1, 0, 1, \dots, L-1, L\}$.

This shows that through an increase in the number of encoding repetitions in the QAUM circuit, it is possible to approximate the classification problem with a truncated Fourier series of higher frequency (Kordzanganeh et al. 2021).

2.2 Classical ensemble techniques

Ensemble methods were developed as a way to combine the predictions of multiple machine learning models in a way that improves the overall performance of the system. It works by utilizing the diversity of the models in the ensemble to create a final prediction that is more robust and accurate than any one model in the ensemble (Dietterich 2000). The two types of ensemble methods that are explored in this work are bagging and boosting (Dietterich 2000).

A bagging ensemble utilizes multiple high accuracy learners and combines their classification outputs using weighted voting to build a powerful ensemble with higher accuracy than any of the constituent classifiers. The training methodology of the ensemble is shown in Fig. 1, where each learner is assigned its own dataset, sampled with replacement from the original training dataset, thereby introducing diversity into the ensemble. Each dataset may contain duplicate data points due to the replacement in the sampling process, and there may also be overlap between the different datasets. This

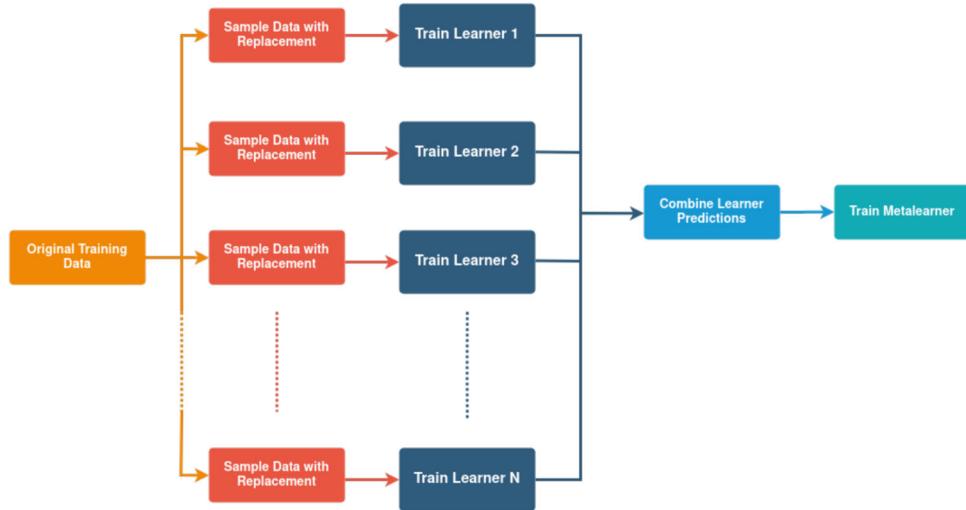


Fig. 1 Here, we describe the procedure used to train the bagging ensemble. The original training dataset is sampled N times, with replacement, to create a unique training dataset for each of the N learners in the ensemble. These learners are then all trained, a process which can be

parallelized due to the mutual independence of the learners. All of the trained learners are then used to classify the training data, and their outputs are combined to create the training data for the metalearner, which is then trained

process helps prevent the ensemble from overfitting the original dataset and is called bootstrap-aggregating (aka bagging). It also allows the learners to be trained in parallel as they are mutually independent.

A boosting ensemble utilizes multiple weak individual learners (with validation accuracies slightly better than 50%) and combines them using weighted voting to build a powerful ensemble. The strength of the ensemble comes from the mechanism through which the datasets each learner is trained on are generated, and is illustrated in Fig. 2. The initial learner is trained on the specified dataset and is used to classify the points in the dataset. Each data point is assigned an equal weight at the start of the training, and thereafter, the weights of the correctly and incorrectly classified data points are decreased and increased, respectively. The dataset is then resampled using the new weights, thereby emphasizing

the datapoints that are difficult to classify correctly. This new dataset is given to the next learner, and this process is repeated until a classifier which is entirely correct or entirely incorrect is obtained.

One of the most influential aspects of ensemble methods is the manner in which the predictions of all of the individual models in the ensemble are combined to produce a single output. One of the most common methods is a weighted voting scheme whereby the outputs of each of the ensemble's constituent models are considered and the degree to which they contribute is decided by the weight assigned to the model (Dietterich 2000). As this choice of weights for each of the models in the ensemble is extremely important, we utilized a metalearner that learns the optimal weights to maximize the performance of the ensemble. We create the labeled training data for the metalearner using both the predictions of the

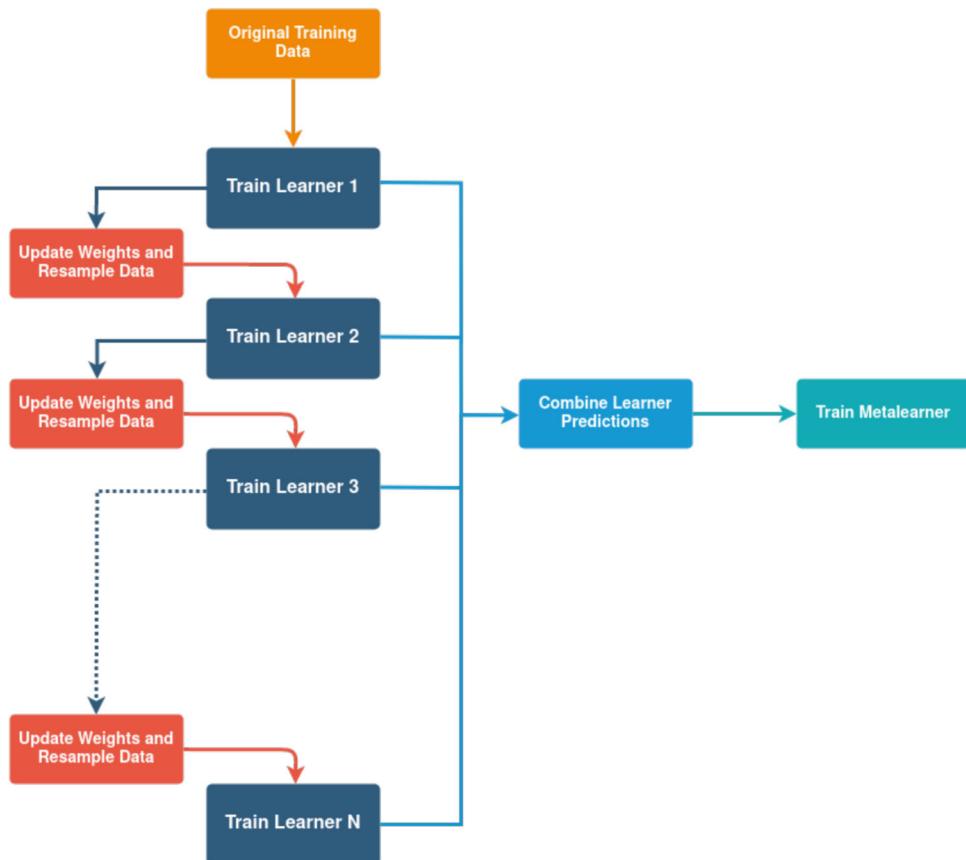


Fig. 2 Here, we describe the procedure used to train the boosting ensemble. The initial learner in the ensemble is trained on the original training data and is used to perform the classification. The weights of the datapoints that are correctly and incorrectly classified are decreased and increased, respectively. This allows the misclassified datapoints to appear more frequently in the subsequent samples of the training data.

After resampling with the new weights, the next learner is trained on this new dataset. This process is repeated until either the limit on ensemble size is reached or until a learner is either entirely correct or incorrect. Next, all of the learners in the ensemble are used to classify the training data, and their outputs are combined to create the training data for the metalearner, which is then trained

individual models on the original training data and the true class labels of the data points. Once trained, the metalearner is capable of combining the predictions of all of the models to produce the most accurate ensemble possible from the data. In this work, we select a logistic regression model to use as our metalearner.

Despite ensemble methods offering a host of advantages, such as improved robustness, accuracy, and generalization capabilities, they are not without faults (Goodfellow et al. 2016). The use of multiple learners in the ensemble increases the complexity and computational cost of training the ensemble (Goodfellow et al. 2016). Additionally, there are a myriad of base learners to choose from, each with their own benefits and drawbacks, so choosing the makeup of the ensemble or the hyper-parameters of the learners in the ensemble is no simple matter (Goodfellow et al. 2016).

3 Methodology

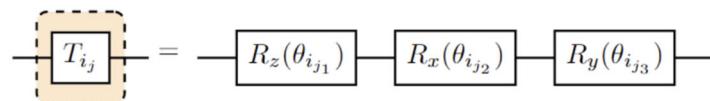
3.1 Quantum classifier

The selection of the individual classifiers that make up the ensemble has a strong effect on the overall success of the ensemble techniques (Dietterich 2000). In this work, the ensembles consist of QAUM circuits, whose design was pro-

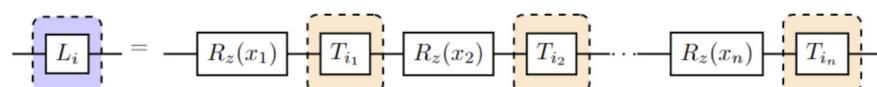
posed in Kordzanganeh et al. (2021). QAUM circuits, seen in Fig. 3, utilize only a single qubit and are composed of layers of parameterized Z , X , and Y -rotation gates. These rotations make up the trainable blocks and are interlaced with Z -rotation gates to encode the data features (Kordzanganeh et al. 2021). This choice of gates allows efficient access to both the group and Fourier spaces, and the first degree multi-dimensional Fourier series of the dataset can be expressed by this architecture (Kordzanganeh et al. 2021).

In Schuld et al. (2021), it was shown that quantum models can naturally be expressed as a partial Fourier series in the data and that the accessible spectrum of frequencies is dictated by the nature of the data encoding gates utilized in the circuit. Once a quantum model has access to a sufficiently large spectrum of frequencies, it becomes a universal function approximator for periodic functions (Schuld et al. 2021).

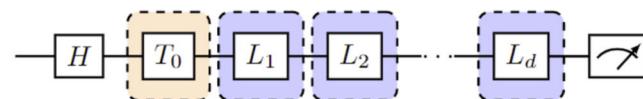
The choice of rotation gates in the trainable layers used in this work was modified so that the optimizer can more easily train the parameters for these gates. The original structure of Z , X , and Y -rotations in Fig. 4a was changed to the Z , X , and Z -rotations shown in Fig. 4b. As the Y -rotation in the original circuit is composed of both an X and Z -rotation, changing it to a Z -rotation allows the optimizer to more easily train the parameters of this gate to minimize the loss function as it reduces the complexity of cost function that the optimizer is minimizing. This can facilitate a more effective search for



(a) A trainable block in the QAUM circuit contains sequential R_z , R_x , and R_y gates.



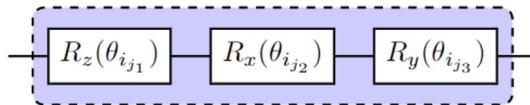
(b) A layer in the QAUM circuit consists of alternating R_z encoding gates and trainable blocks.



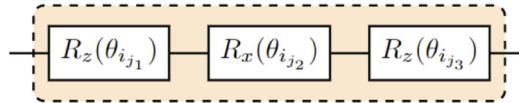
(c) A full QAUM circuit of depth $= d$. The depth is variable and can be selected to suit the application.

Fig. 3 The QAUM circuit from Kordzanganeh et al. (2021) utilizes an alternating structure of single-qubit data encoding gates and trainable layers of parameterized single-qubit gates. This structure facilitates an

increase in model expressivity by increasing the number of layers in the circuit. The structure of the trainable blocks, circuit layers, and complete QAUM circuit can be seen in **a**, **b**, and **c**, respectively



(a) The original gate selection for the trainable layers devised in Kordzanganeh et al (2021).



(b) The modified gate selection used for the trainable layers in this work.

Fig. 4 In this work, we modify the gate selection in the trainable layers of the QAUM circuit from Kordzanganeh et al. (2021). By utilizing the ZXZ-rotations in **b**, rather than the ZXY-rotations in **a**, the optimizer is able to more easily converge on the optimal solution thanks to a sim-

pler gate composition. The trainable layer shown in **b** is still capable of efficiently accessing the Fourier space as it can turn any initial state into any final state (Nielsen and Chuang 2002)

the optimal parameters for the classifiers (Goodfellow et al. 2016). Additionally, this modification has no effect on the accessibility of the group and Fourier spaces as the rotation ZXZ can still convert any initial state $|\psi\rangle$ into any final state $|\phi\rangle$ (Nielsen and Chuang 2002).

3.2 Datasets

The main dataset used in this work is the HTRU 2 Pulsar dataset (Lyon et al. 2016). This is the binary classification of 8-dimensional input as either radio frequency interference (RFI) or a pulsar star, an important problem for the analysis of gravitational waves and other applications (Kordzanganeh et al. 2021).

To test whether the performance trends for the classifiers are consistent across different problems, we also utilized a subset of MNIST Digits (LeCun and Cortes 2010) containing

the digits 8 and 9, and the Wisconsin Breast Cancer dataset (Dua and Graff 2017). In all cases, we tested using only 8-dimensional input, so principal component analysis was used to perform dimensionality reduction where needed.

For the purposes of result validation, we conducted a paired *t*-test to compare the classification models. We performed 5×2 cross-validation on the datasets, whereby 5 different samples of the dataset are taken, shuffled, and then split for use in training and validation for each of the models.

3.3 Optimization and implementation

The training of the quantum models which make up the ensembles was performed using the constrained optimization by linear approximation (COBYLA) optimizer (Powell 1998), a gradient-free optimizer which serves our purposes well as the quantum circuits are executed on a noise-free clas-

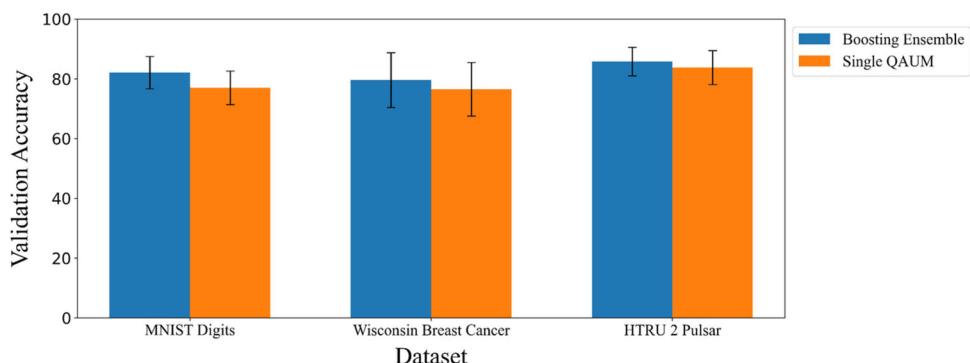


Fig. 5 The mean validation accuracies of the boosting ensemble (blue) and original QAUM classifier (orange) are shown, with error bars indicating the standard deviation of the accuracies. Testing was performed using 5×2 cross-validation for each dataset. We find that the boosting

ensemble outperforms the original QAUM classifier for the HTRU 2 Pulsar dataset (Lyon et al. 2016), as well as the MNIST Digits and Wisconsin Breast Cancer datasets (LeCun and Cortes 2010; Dua and Graff 2017)

Table 1 The classifier details and validation accuracies for the testing performed on the HTRU 2 Pulsar, MNIST Digits, and Wisconsin Breast Cancer datasets used (Lyon et al. 2016; LeCun and Cortes 2010; Dua and Graff 2017)

Validation accuracy and training time for each dataset					
Type of classifier	Number of learners	Circuit depth	Epochs per learner	Digits dataset accuracy	Cancer dataset accuracy
Boosting ensemble	27	1	1	82.1 ± 5.4	79.6 ± 9.2
Single QAUM	1	2	100	77.0 ± 5.6	76.5 ± 9.0

The number of learners for the boosting ensemble is taken as the average over all 3 datasets, as the ensemble determines this number based on the difficulty of the dataset. The boosting ensemble achieves higher classification accuracies, with generally lower standard deviations, than the single QAUM classifier in each of the classification problems and is more than twice as fast to train

sical simulator. In all testing cases, the termination criteria for the optimizer was specified to be a maximum number of function evaluations, which will be referred to as the total number of training iterations in the next section.

The implementation of all of the ensembles, as well as the original QAUM classifier from Kordzanganeh et al. (2021), can be found at McFarthing (2022). This was accomplished using the Julia programming language (Bezanson et al. 2017), and the Yao (Luo et al. 2020) library was used for the creation and execution of the quantum variational circuits making up the ensembles' constituent models.

3.4 Testing

Due to the differences between the bagging and boosting ensembles implemented, specific hyper-parameters, including the number of training iterations per learner and the number of learners in the ensemble, were selected for each. The bagging algorithm works best with accurate base learners, so we trained the learners for a large number of iterations and used only a few learners in the ensemble to keep training overhead down. We tested a bagging ensemble of 7 QAUM classifiers, each with 2 embedding layers, and each learner was trained for 1714 iterations. This ensemble was then compared to a single QAUM classifier, also with 2 embedding layers, trained for 12,000 iterations. The number of iterations per learner was selected by first choosing the amount of training for the single QAUM classifier and then splitting the number of iterations evenly among the learners in the ensemble.

On the other hand, the boosting algorithm relies on learners which do not have high accuracies individually, so the number of embedding layers in the quantum circuit was set

to 1 and each learner was only trained for 1 iteration. This ensemble was then compared to a single QAUM classifier, with 2 embedding layers, trained for 100 iterations. Though it is not known how many learners the boosting ensemble will need to reach convergence prior to executing the algorithm, the maximum number of learners was limited to 100 and the number of iterations per learner was set to 1. Thus, the ensemble would not go through more training iterations than the single QAUM classifier, even if it did not reach convergence.

4 Results

To verify the significance of the results obtained from our testing, a paired *t*-test with a threshold of 0.05 was utilized in conjunction with 5×2 cross-validation. In all tests, the *p*-value was significantly lower than the threshold, thereby confirming the legitimacy of the results.

Figure 5 and Table 1 show that the boosting ensemble outperformed the single QAUM classifier for each of the datasets used. The boosting ensemble had validation accuracies that were 2%, 5.1%, and 3.1% higher than those of the single QAUM classifier for the Pulsar, Digits, and Cancer datasets, respectively. Additionally, the time taken to train the boosting ensemble was less than half that of the single QAUM classifier. The validation accuracies for each of the samples in the 5×2 cross-validation can be found in Appendix A.

Figure 6 and Table 2 show that the bagging ensemble outperformed the single QAUM classifier for each of the datasets used. The bagging ensemble had validation accuracies that were 1.6%, 3.1%, and 0.6% higher than those of the single QAUM classifier for the Pulsar, Digits, and Cancer datasets,

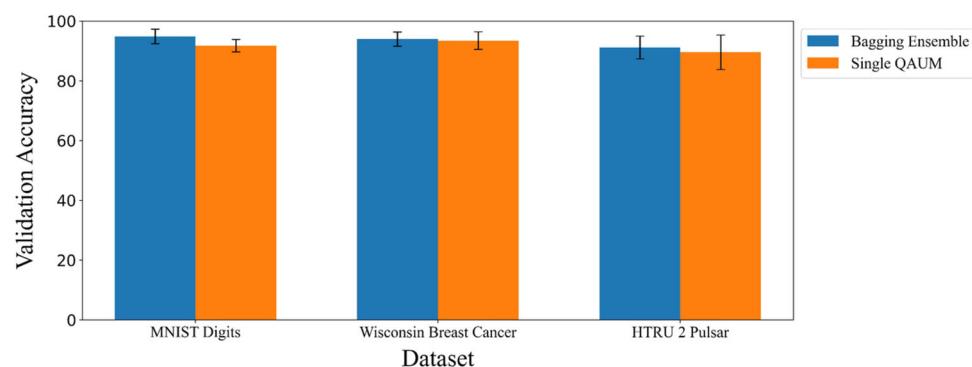


Fig. 6 The mean validation accuracies of the bagging ensemble (blue) and original QAUM classifier (orange) are shown, with error bars indicating the standard deviation of the accuracies. Testing was performed using 5×2 cross-validation for each dataset. We find that the bagging

ensemble outperforms the original QAUM classifier for the HTRU 2 Pulsar dataset (Lyon et al. 2016), as well as the MNIST Digits and Wisconsin Breast Cancer datasets (LeCun and Cortes 2010; Dua and Graff 2017)

Table 2 The classifier details and validation accuracies for the testing performed on the HTRU 2 Pulsar, MNIST Digits, and Wisconsin Breast Cancer datasets used (Lyon et al. 2016; LeCun and Cortes 2010; Dua and Graff 2017)

Validation accuracy and training time for each dataset						
Type of classifier	Number of learners	Circuit depth	Epochs per learner	Digits dataset accuracy	Cancer dataset accuracy	Pulsar dataset accuracy
Bagging ensemble	7	2	1714	94.9 ± 2.4	94.0 ± 2.4	91.2 ± 3.8
Single QAUM	1	2	12000	91.8 ± 2.1	93.4 ± 2.9	89.6 ± 5.8

The bagging ensemble achieves higher classification accuracies, with generally lower standard deviations, than the single QAUM classifier in each of the classification problems. Similarly to the boosting ensemble, the bagging ensemble is also quicker to train than the comparative QAUM classifier

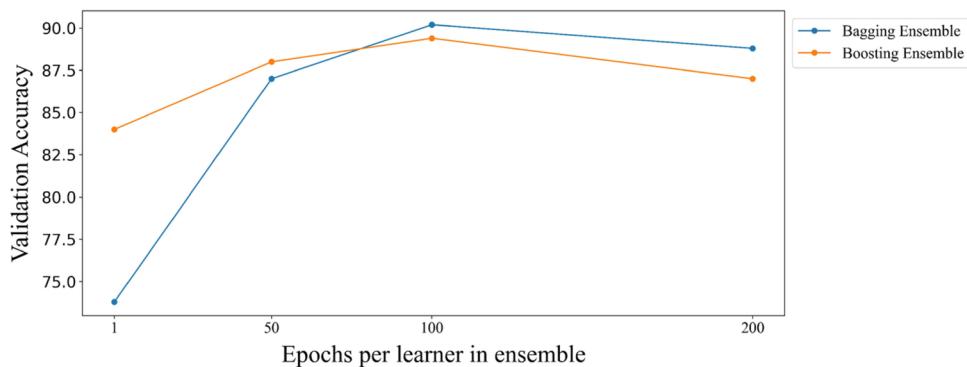


Fig. 7 Validation accuracy of the bagging and boosting ensembles on the HTRU 2 Pulsar (Lyon et al. 2016) dataset with 100 samples. The trade-off between training time requirements and overall accuracy for the two types of ensembles can be seen

respectively. The validation accuracies for each of the samples in the 5×2 cross-validation can be found in Appendix B.

Tables 1 and 2 both show that in most cases, the accuracies from the ensemble classifiers have a lower standard deviation than those from the single QAUM classifier. This appears to show that in general, the bagging and boosting ensembles have more consistent performance over the samples of data from each dataset. This is desirable as it indicates that the ensembles may be less sensitive to variations in the input data.

It is noticeable that the accuracies achieved by the boosting ensemble in Table 1 are lower than those of the bagging ensemble in Table 2. Additionally, the duration of training was far lower for the boosting ensemble than for the bagging ensemble, thus raising the question of the use case for each of these ensembles.

In Fig. 7, the validation accuracies for both types of ensembles are shown in relation to training length per learner, and it is clear that the boosting ensemble is able to achieve a decent level of accuracy with almost no training at all. However, as the number of training epochs is increased, the bagging ensemble overtakes the boosting ensemble for higher overall accuracy. Our analysis involves the training of both types of ensembles to 200 iterations per learner, as it can be seen that after about 150 iterations, the accuracies of the ensembles begin to degrade. This appears to signify that they start to overfit the training data, resulting in poorer generalization capabilities for both ensembles.

5 Conclusions

In this work, we explored the use of classical ensemble techniques, specifically bagging and boosting, as a means to expand the classification power of the QAUM classifier from Kordzanganeh et al. (2021).

Through testing on multiple datasets (Lyon et al. 2016; LeCun and Cortes 2010; Dua and Graff 2017), we have shown

that both boosting and bagging ensembles outperform the original QAUM circuit. This is shown in Figs. 5 and 6. Additionally, we analyzed the differences between the results of the boosting and bagging ensemble, shown in Fig. 7, and propose that the boosting ensemble of QAUM classifiers is applicable to problems where training time is prioritized over accuracy. This is due to its ability to achieve a reasonable level of performance very quickly. Furthermore, we propose that the bagging ensemble be used when accuracy is of the utmost performance as it is capable of generalizing better than the boosting ensemble with sufficient training.

The application of these classical ensemble techniques to the QAUM single-qubit circuits was investigated as an alternative to other multi-qubit approaches, and the results obtained in this work show that this is promising as improvements in classification accuracy can be achieved with relatively shallow circuits. The original QAUM circuit (Kordzanganeh et al. 2021) was designed to achieve higher accuracy with increasing circuit depth. However, due to the low coherence time of current devices, this would not be practical. Through the use of the techniques in this work, a powerful ensemble can be built from shallow single-qubit circuits which can be run on existing quantum hardware.

The testing of this approach on real quantum hardware can be done in future research, as well as investigations regarding the noise resilience of these shallow ensembles compared to the original QAUM classifier.

A. Cross-validation results for boosting ensemble vs. single QAUM classifier

The validation accuracies for each of the dataset samples used in the cross-validation are shown in Figs. 8, 9, and 10. As mentioned previously, the 3 datasets used in testing were HTRU 2 Pulsar, MNIST Digits, and Wisconsin Breast Cancer (Lyon et al. 2016; LeCun and Cortes 2010; Dua and Graff 2017).

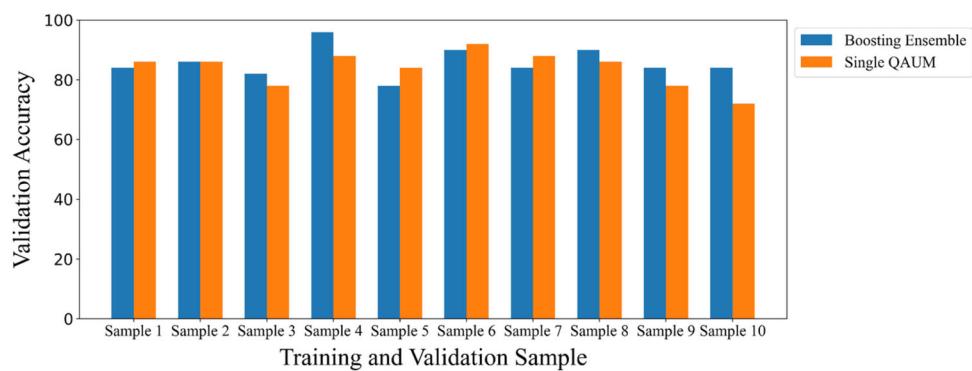


Fig. 8 Validation accuracies of the boosting ensemble (blue) and single QAUM classifier (orange) on the HTRU 2 Pulsar (Lyon et al. 2016) dataset for each of the samples used in the 5×2 cross-validation

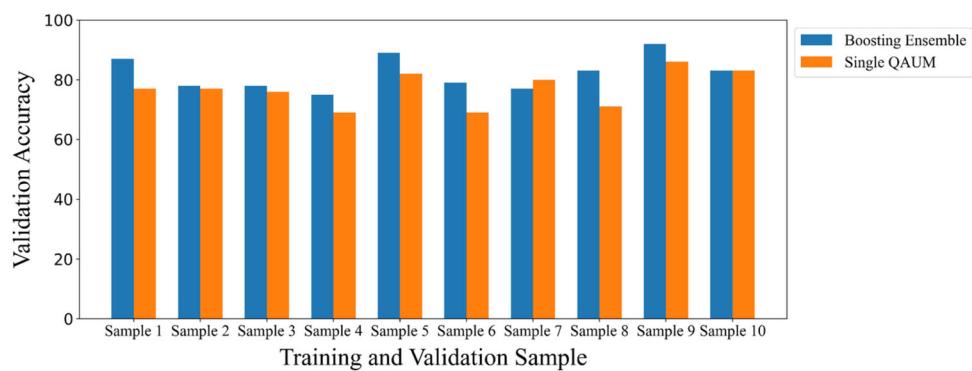


Fig. 9 Validation accuracies of the boosting ensemble (blue) and single QAUM classifier (orange) on the MNIST Digits (LeCun and Cortes 2010) dataset for each of the samples used in the 5×2 cross-validation

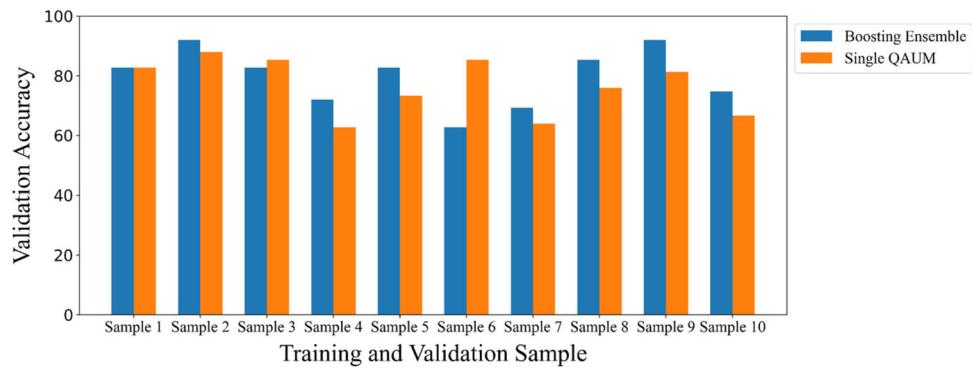


Fig. 10 Validation accuracies of the boosting ensemble (blue) and single QAUM classifier (orange) on the Wisconsin Breast Cancer (Dua and Graff 2017) dataset for each of the samples used in the 5×2 cross-validation

B. Cross-validation graphs for bagging ensemble vs. single QAUM classifier

The validation accuracies for each of the dataset samples used in the cross-validation are shown in Figs. 11, 12, and

13. As mentioned previously, the 3 datasets used in testing were HTRU 2 Pulsar, MNIST Digits, and Wisconsin Breast Cancer (Lyon et al. 2016; LeCun and Cortes 2010; Dua and Graff 2017).

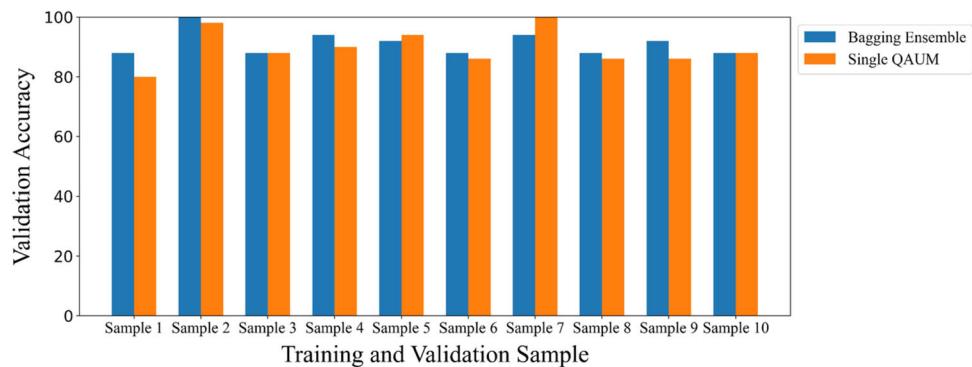


Fig. 11 Validation accuracies of the bagging ensemble (blue) and single QAUM classifier (orange) on the HTRU 2 Pulsar (Lyon et al. 2016) dataset for each of the samples used in the 5×2 cross-validation

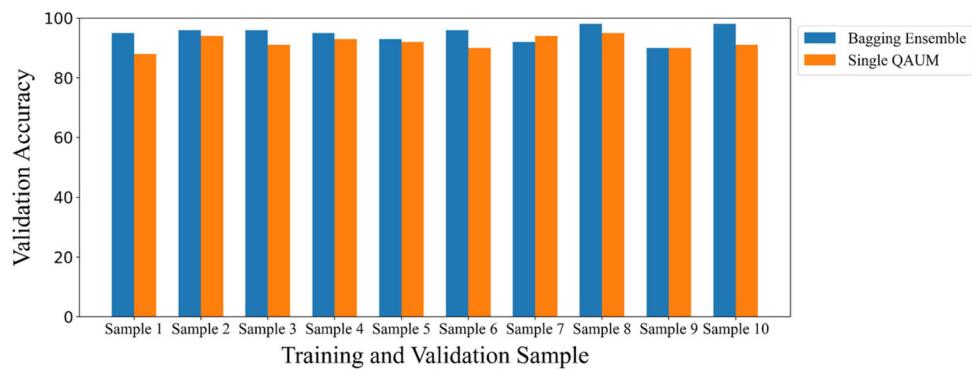


Fig. 12 Validation accuracies of the bagging ensemble (blue) and single QAUM classifier (orange) on the MNIST Digits (LeCun and Cortes 2010) dataset for each of the samples used in the 5×2 cross-validation

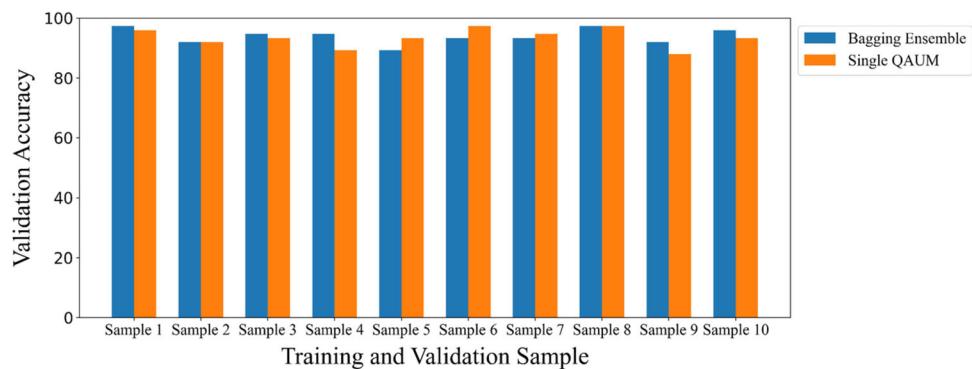


Fig. 13 Validation accuracies of the bagging ensemble (blue) and single QAUM classifier (orange) on the Wisconsin Breast Cancer (Dua and Graff 2017) dataset for each of the samples used in the 5×2 cross-validation

Author Contributions “SMF wrote the main manuscript text. All authors reviewed the manuscript.”

Funding Open access funding provided by University of KwaZulu-Natal. We would like to thank the National Institute of Theoretical and Computational Sciences (NITheCS) and the Center for Artificial Intelligence Research (CAIR) for funding this research.

Availability of Data and Materials All code and data associated with the current submission are available at <https://github.com/shanemcfarthing/Classical-Ensembles-of-Single-Qubit-Quantum-Classifiers>.

Declarations

Competing Interests Francesco Petruccione is the Chair of Scientific Board and Co-Founder of QUNOVA computing. The authors declare no other competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Bezanson J, Edelman A, Karpinski S et al (2017) Julia: a fresh approach to numerical computing. *SIAM review* 59(1):65–98. <https://doi.org/10.1137/141000671>

Dietterich TG (2000) Ensemble methods in machine learning. In: International workshop on multiple classifier systems, Springer, pp 1–15. https://doi.org/10.1007/3-540-45014-9_1

Dua D, Graff C (2017) UCI machine learning repository. <http://archive.ics.uci.edu/ml>

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press

Kordzanganeh M, Utting A, Scaife A (2021) Quantum machine learning for radio astronomy. <https://doi.org/10.48550/arXiv.2112.02655>

LeCun Y, Cortes C (2010) MNIST handwritten digit database. <http://yann.lecun.com/exdb/mnist/>

Li Q, Huang Y, Hou X, et al. (2023) Ensemble-learning variational shallow-circuit quantum classifiers. <https://doi.org/10.48550/arXiv.2301.12707>

Luo XZ, Liu JG, Zhang P, et al (2020) Yao. jl: extensible, efficient framework for quantum algorithm design. *Quantum* 4:341 <https://doi.org/10.22331/q-2020-10-11-341>

Lyon RJ, Stappers B, Cooper S et al (2016) Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Month Notice Royal Astronom Soc* 459(1):1104–1123. <https://doi.org/10.1093/mnras/stw656>

Macaluso A, Clissa L, Lodi S, et al. (2020) Quantum ensemble for classification. <https://doi.org/10.48550/arXiv.2007.01028>

McClean JR, Boixo S, Smelyanskiy VN et al (2018) Barren plateaus in quantum neural network training landscapes. *Nat Commu* 9(1):1–6. <https://doi.org/10.1038/s41467-018-07090-4>

McFarthing S (2022) Classical ensembles of single-qubit quantum variational circuits for classification. <https://github.com/shanemcfarthing/Classical-Ensembles-of-Single-Qubit-Quantum-Classifiers.git>

Nielsen MA, Chuang I (2002) Quantum computation and quantum information. American Ass Physics Teacher <https://doi.org/10.1119/1.1463744>

Ortiz Marrero C, Kieferová M, Wiebe N (2021) Entanglement-induced barren plateaus. *PRX Quantum* 2. <https://doi.org/10.1103/PRXQuantum.2.040316>

Powell MJ (1998) Direct search algorithms for optimization calculations. *Acta Numer* 7:287–336. <https://doi.org/10.1017/S0962492900002841>

Qin R, Liang Z, Cheng J, et al (2022) Improving quantum classifier performance in NISQ computers by voting strategy from ensemble learning. <https://doi.org/10.48550/arXiv.2210.01656>

Schuld M, Petruccione F (2018) Quantum ensembles of quantum classifiers. *Sci Reports* 8(1):1–12. <https://doi.org/10.1038/s41598-018-20403-3>

Schuld M, Petruccione F (2018) Supervised learning with quantum computers, vol 17. Springer. <https://doi.org/10.1007/978-3-319-96424-9>

Schuld M, Sinayskiy I, Petruccione F (2015) An introduction to quantum machine learning. *Contemp Physics* 56(2):172–185. <https://doi.org/10.1080/00107514.2014.964942>

Schuld M, Sweke R, Meyer JJ (2021) Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Phys Rev A* 103(3):032,430. <https://doi.org/10.1103/PhysRevA.103.032430>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.